

A Quantum Measurement Scenario which Requires Exponential Classical Communication for Simulation

Allison Coates
Computer Science Division
University of California at Berkeley
allisonc@cs.berkeley.edu

Abstract

In this paper we consider the following question: how many bits of classical communication and shared random bits are necessary to simulate a quantum protocol involving Alice and Bob where they share k entangled quantum bits and do not communicate at all. We prove that 2^k classical bits are necessary, even if the classical protocol is allowed an ϵ chance of failure.

1. Introduction

Quantifying the power of shared entanglement is a fundamental goal of quantum information theory. In this paper, we compare the power of shared entanglement to the power of classical correlations with communication by considering the following restricted type of quantum protocol, which we call **sampling-using-entanglement**: Alice and Bob are two parties which are not allowed to communicate with each other, but they share k entangled quantum bits. Each party makes a generalized measurement on his quantum bits.

We compare this protocol to a classical simulation of it, where Alice and Bob use classical correlations while communicating classical bits to sample from a distribution that is ϵ -close of the distribution resulting from the quantum protocol.

The cost of the classical simulation is the sum of the number of shared random bits and the number of classical bits communicated between the two parties. The number of shared random bits is the classical analogue to the amount of shared entanglement, while classical communication serves to augment the limited power of classical correlations.

Our main result is that 2^k classical bits are needed to simulate the distribution that arises from the quantum protocol's measurement on k entangled qubits. Our re-

sult is based on the analysis of sampling protocols for the set-disjointness problem $\text{DISJ}(x,y)$ [?]. In that paper, the authors demonstrate that the set-disjointness problem can be sampled using $O(\log(n) \log(1/\epsilon))$ qubits of communication between two parties, while classical sampling of set-disjointness requires $\Theta(\sqrt{n})$ classical bits, whether those bits are communicated online during the protocol, or are a cache of prior correlations. We cast these results in a different light, showing how this result holds for a protocol in which the parties share entangled bits. In particular, we demonstrate an efficient algorithm for sampling the set-disjointness function $\text{DISJ}(x,y)$ using $O(\log(n) \log(1/\epsilon))$ entangled bits and no other communication, which is again an exponential improvement over the classical sampling complexity of set-disjointness. These results show that there exist problems for which classical simulation of an entangled state would require an exponential amount of communication, demonstrating that classical correlations are less powerful than quantum correlations.

Another intriguing element of this result is that in this quantum protocol each party maintains some secret information about their respective subset, so that parties do not have full information about the other parties' output. Such a protocol is fundamentally different than known classical algorithms, where cryptographic schemes are necessary for A and B to stay ignorant about each other's outputs.

Within the community of quantum communication complexity research, there is growing body of work addressing the cost of simulation of entanglement using classical bits[?], [?], [?], [?], [?], and similar work in the communication complexity of entanglement [?]. The work of Brassard et. al. [?] showed that 8 classical bits of communication suffice to exactly simulate a Von Neumann measurement on a Bell pair if an infinite random string is shared between two parties. That paper also shows a partial function for which an exact classical simulation of an entangled system of n qubits

would require 2^n classical bits of communication. However, they left open the question of whether any classical simulation existed on a more robust quantum measurement scenario when that simulation allowed small error, or was limited in the amount of bits available in prior correlation. Massar et. al. [?] showed that no protocol restricted to a finite amount of prior correlation could simulate the Bell pair, but that Bell states can be simulated without an infinite random string if the expected classical communication is finite.

Our work differs from the models studied above because we consider the the communication complexity of sampling rather than of functions or partial functions. Further, we demonstrate a protocol that is robust to small errors. Lastly, as shown in Massar et. al. [?] an infinite amount of shared randomness is too powerful to be used as a resource. As such, we quantify the total amount of classical information that is used in the protocol, counting the amount of prior correlations as well as the amount of communicated bits. Finally, we believe that the sampling model is more appropriate for investigation of quantum communication complexity problems since measurement of a quantum mechanical system is inherently a sampling process.

In 2 we discuss classical and quantum sampling protocols and introduce our restricted quantum protocol for sampling-using-entanglement. In 4 we state the problem of set-disjointness, and introduce an sampling-with-entanglement protocol for it. Section 5 contains the main theorem proof of the upper bound on the communication complexity of sampling-using-entanglement for the set-disjointness problem, while 8 presents the proof of the main lemma for this result. In 6 we present the classical sampling results for set-disjointness, demonstrating an exponential gap between the communication complexity of sampling-using-entanglement and communication complexity of classical sampling.

2. Preliminaries

Consider a scenario in which two parties, Alice and Bob wish to output samples according to some known joint probability distribution. They share $|\psi_{AB}\rangle$, a bipartite entangled state of k quantum bits, but they cannot communicate classically with each other. A simple protocol, labelled $\{U_A, U_B, |\psi_{AB}\rangle\}$, samples a probability distribution as follows: each party performs a unitary operation, U_A and U_B , respectively, on their portions of the entangled system $|\psi_{AB}\rangle$, and additional ancilla. Then, each party performs a measurement in the computational basis and outputs the results. Since a measurement of a quantum system re-

sults in a classical probability distribution, this protocol samples from the classical probability distribution $\Pi\{U_A, U_B, |\psi_{AB}\rangle\}$ induced by performing a measurement $U_A \otimes U_B$ on $|\psi_{AB}\rangle$. For shorthand, we will refer to a unitary operation and subsequent measurement in the computational basis as a **generalized measurement**, and say that the protocol samples a state $(U_A \otimes U_B) |\psi_{AB}\rangle$ to mean that the protocol outputs samples according to the classical probability distribution $\Pi\{U_A, U_B, |\psi_{AB}\rangle\}$.

Now imagine Alice and Bob collude; that is, they claim to share an entangled state, but in reality they have only classical communication and prior classical correlations available to them. They will again be given unitary operators and asked to output measurements according to $\Pi\{U_A, U_B, |\psi_{AB}\rangle\}$. In order for them to succeed in convincing a third party that they share entanglement, they must “simulate” this quantum protocol by outputting results according to a probability distribution that is ϵ -close to the probability distribution $\Pi\{U_A, U_B, |\psi_{AB}\rangle\}$.

We now define formally a two-party quantum protocol for the sampling-using-entanglement of $|\psi_{AB}\rangle$ and the subsequent classical simulation of such a protocol. To do so, we include an intermediate definition of q-sampling [?] where the protocol that samples $|\psi_{AB}\rangle$ not only outputs values according to the distribution of $|\psi_{AB}\rangle$ but creates a pure state that is close to the $|\psi_{AB}\rangle$ desired.

Definition 1 *Let $|\psi_{AB}\rangle$ be a bipartite entangled state, and let U_A and U_B be unitary operators. A two-party **sampling-using-entanglement protocol** Q is a quantum protocol executed by Alice and Bob as follows: Alice and Bob perform their respective unitary operators on their portions of the $|\psi_{AB}\rangle$ and some additional ancilla. Then, each party performs a generalized measurement in the computational basis. The two parties do not communicate in any other way. This protocol outputs samples according to the classical probability distribution $\Pi\{U_A, U_B, |\psi_{AB}\rangle\}$, defined as probability distribution induced by performing a measurement $U_A \otimes U_B$ on $|\psi_{AB}\rangle$. Further, we say that Q **q-samples** a state $|\Psi\rangle$ if $\langle\Psi|(U_A \otimes U_B)|\psi_{AB}\rangle > 1 - \epsilon$. The communication complexity of sampling-using-entanglement is the number of quantum bits shared between the two parties.*

Definition 2 *Let Q be a sampling-using-entanglement protocol as above that samples the distribution $\Pi\{U_A, U_B, |\psi_{AB}\rangle\}$. Let P be a two-party classical protocol P where each party is computationally unbounded. We say P is a **classical simulation** of Q if P outputs samples x and y according to a distribution $\Pi'(P)$, such*

that that $|\Pi'(P) - \Pi\{U_A, U_B, |\psi\rangle\}|_{TVD} < \epsilon$, where the norm is the Total Variational Distance. The communication complexity of P is $k + m$, the sum of the number of classical bits of communication (k) and the number of classical bits of correlation (m).

The classical simulation does not need to simulate the specific quantum state or its overall phase information; it only matches the overall probability distribution. However, the quantum sampling protocol we describe is more restrictive, and matches the appropriate quantum state, as well as the resulting distribution.

In previous literature [?], the communication complexity of classical sampling is defined as the number of classical bits of communication needed to sample a function. This definition is equivalent to the number of bits of correlation needed to sample f .

3. Discussion

The main result of this paper is to answer the question of whether all quantum measurement scenarios can be simulated with only a polynomial number of classical bits. We demonstrate that the answer is no, by showing that already established results in quantum communication complexity can be applied directly to this problem and provide the answer.

The main technical theorems of this paper were first proved by Ambainis et. al. [?] in a paper comparing the classical and quantum communication complexities of sampling. While we provide the details of those theorems again for completeness, our main result is to show that those results apply directly to a new communication protocol of sampling-using-entanglement. A sampling-using-entanglement protocol is equivalent to a quantum measurement scenario: both protocols involve each party performing a generalized measurement on their portion of the entangled state.

Prior work in quantum information theory has failed to answer the question of whether all larger non-separable entangled systems could be simulated with even a polynomial number of bits. However, by applying the communication complexity results for sampling, we find that if such a scenario could be simulated with less than an exponential number of bits relative to the quantum scheme, then we could beat the classical sampling bound proven in Ambainis et. al. Therefore, any classical simulation of such a quantum measurement scenario must match the classical communication lower bound on sampling. We use the results of Ambainis et. al. to demonstrate a quantum measurement scenario for the set-disjointness problem on n entangled qubits cannot be simulated with less than 2^n classical bits of communication or correlation.

4. Results for sampling of set-disjointness

We now show that there exists a sampling-using-entanglement quantum protocol for the set-disjointness problem that uses only $O(\log n \log 1/\epsilon)$ qubits. This result provides an exponential speedup over the classical sampling result for set-disjointness, where $\Theta(\sqrt{n})$ classical bits of communication and shared randomness are required. The treatment provided here is based on the treatment in Ambainis et. al, but differs greatly in its organization and presentation.

Formally, the problem of sampling set-disjointness (for size \sqrt{n}) is defined as follows: Given a set $\Omega = \{1, \dots, n\}$, two parties A and B sample the set-disjointness function if they output uniformly random disjoint subsets S and T respectively, $S, T \in \Omega$, $|S| = |T| = \Theta(\sqrt{n})$. Let $|\chi\rangle$ be the superposition representing all possible disjoint subsets of size \sqrt{n} : $|\chi\rangle = \sum_{i,j \cap \emptyset} |i\rangle|j\rangle$. A protocol can quantum sample $|\chi\rangle$ if it creates a state $|\phi\rangle$ which is close to $|\chi\rangle$. The trivial sampling-using-entanglement protocol quantum samples this function: it simply measures $|\chi\rangle$ in the computational basis. However, the number of qubits necessary to express $|\chi\rangle$ is $\log n(\sqrt{n}) \approx \sqrt{n} \log n$, which does not improve upon the classical sampling complexity of this problem. Instead, we will show that there exists a smaller bipartite system that contains nearly all of the necessary information about $|\chi\rangle$. Specifically, we will construct a pure state $|\psi\rangle$ of only $O(\log(n) \log(1/\epsilon))$ entangled qubits which when operated upon by local unitary operations U_A and U_B yields a state which is an ϵ -approximation for $|\chi\rangle$.

Theorem 1 *There exists a basis V for $|\chi\rangle$, the superposition ranging over all possible disjoint subsets of size \sqrt{n} with the following property: Let $|\psi\rangle$ be the projection of $|\chi\rangle$ onto the subspace spanned by the largest $n^{\log 1/\epsilon}$ eigenvectors of V . Then $\langle \chi | (V \otimes V) | \psi \rangle > 1 - \epsilon$.*

5. Proof of Main Theorem

Proof: We begin by using a matrix representation for states and the unitary operators. Consider $|\chi\rangle$ as above, written as

$$|\chi\rangle = \sum_{i,j} m_{ij} |i\rangle|j\rangle \text{ where } m_{ij} = \begin{cases} 1 & : i \cap j = \emptyset, \\ 0 & : i \cap j \neq \emptyset \end{cases}$$

Let M_χ be the linear operator corresponding to the matrix representation for $|\chi\rangle$: $M_\chi = [m_{ij}]$, where M_χ 's rows are indexed by $|i\rangle$ and columns by $|j\rangle$.

Remark 1 Given a state $|\phi\rangle = \sum_{i,j} m_{i,j} |i\rangle |j\rangle$, a matrix representation $M_\phi = [m_{i,j}]$, and unitary operators U_A and U_B , the matrix representation for the state $(U_A \otimes U_B) |\phi\rangle$ is $U_A M_\phi U_B^\dagger$.¹

Using the representation, a rotation of a state $|\psi\rangle$ into the V basis can be written as $VM_\psi V^\dagger$. This notation is particularly useful because it allows us to compare the norms of pure states by comparing the trace norms of their matrix representations. We define the trace norm of a matrix M as $\|M\| = \sqrt{\text{Tr}(M^\dagger M)}$.

Remark 2 Given states $|\psi\rangle$ and $|\chi\rangle$, their corresponding matrix representations M_ψ and M_χ , and a basis V , then for every ϵ , if $\|VM_\psi V^\dagger - M_\chi\|^2 < 2\epsilon$, then $\langle\chi|(V \otimes V)|\psi\rangle > 1 - \epsilon$.²

With the above two facts in place, it remains to prove that there exists an appropriately-sized basis for the matrices representing our bipartite quantum states.

Lemma 1 *There exists a basis $V = \{v_1, \dots, v_n\}$ for M_χ such that the linear operator M_ψ , the projection of M_χ onto the subspace spanned by $n^{\log(1/\epsilon)}$ eigenvectors of V corresponding to the largest $n^{\log(1/\epsilon)}$ eigenvalues, is within 2ϵ of M_χ :*

$$\|VM_\psi V^\dagger - M_\chi\|^2 < 2\epsilon$$

The proof then follows. ■

Intuitively, this proof follows from the fact that we need only a small number of basis vectors to closely approximate any vector in the vector space of the linear operator M_χ . We can then rotate that system of basis vectors back into a larger space and achieve a linear operator that is close to M_χ . In subsequent sections we prove the associated lemma.

6. Classical Bound comparison

We now return to the issue of simulation of quantum measurement. Since the sampling-using-entanglement protocol is equivalent to a quantum measurement scenario that samples set-disjointness, any classical simulation of this quantum measurement scenario would need to classically sample set-disjointness. But by Ambainis et. al. [?], the lower bound for classical sampling of set-disjointness is $\Omega(\sqrt{n})$. Therefore, any classical protocol simulating the the quantum measurement scenario of set-disjointness must use $\Omega(\sqrt{n})$ bits of communication. This demonstrates an exponential gap between sampling-using-entanglement and classical sampling, and shows that the result of Brassard et. al.

¹To see the proof of this remark, please refer to the Appendix.

²To see the proof of this remark, please refer to the Appendix.

[?] that a Von Neumann measurement on a Bell pair could be simulated with a constant number of classical bits of communication does not scale to larger quantum measurement scenarios on systems of qubits.

7. A cryptographic protocol

This protocol shows that given a small initially entangled state, by performing unitary operations on that state and some ancilla, we can rotate into a larger space. But because the original entangled state resided on a smaller subspace, parties A and B do not have full knowledge of each other's final output. Indeed, they are only aware of $O(\log(n) \log(1/\epsilon))$ bits of information of each others output, rather than \sqrt{n} bits. This is a fundamental difference over classical information theory, where the existence of cryptographic one-way functions is necessary to achieve the same "hidden" bits in a sampling protocol.

We believe that this amount of hidden information could be used to create future cryptographic protocols for property testing. Those result will be shown in a later version of this paper. We also believe that this imbalance between hidden information and classical simulation results is an interesting area for future research into the power of entanglement. This idea addresses some of the concerns expressed in Collins and Popescu [?], who raised the question of what particular behavior of entanglement is quantum-mechanical. We believe that the hidden bits here may directly answer this question.

8. Proof of Lemma 1

Proof: We begin by assuming that such a basis exists, and consider the t eigenvectors that correspond to the largest t eigenvalues. We derive conditions on the value of t that succeeds in achieving an approximation bound of 2ϵ .

Let V be a basis of eigenvectors for M_χ : $M_\chi = V\Lambda V^\dagger$, where Λ is the matrix of eigenvalues. Let V_t denote the subspace defined by the first t eigenvectors of V . We define M' as the linear operator created by projecting M onto the subspace of the first k eigenvectors in V .

$$\begin{aligned}
\|VM'V^\dagger - M_\chi\|^2 &= \|VM'V^\dagger - V\Lambda V^\dagger\|^2 \\
&= \|V(M' - \Lambda)V^\dagger\|^2 \\
&= \|M' - \Lambda\|^2 \\
&= \|V_t M_\chi V_t^\dagger - \Lambda\|^2 \\
&= \sum_{i>t}^{|V|} \lambda_i^2
\end{aligned}$$

Therefore, $\|VM'V^\dagger - M_\chi\|^2 < 2\epsilon$ if $\sum_{i>t}^{|V|} \lambda_i^2 < 2\epsilon$. We continue by characterizing the eigenvalues of M_χ , and use this information to derive a bound on t to ensure $\sum_{i>t}^{|V|} \lambda_i^2 < 2\epsilon$.

M_χ can be written as a linear combination of the ones matrix and a related matrix $B = [b_{ij}]$ of size $N = n()k$ by $N = n()k$, whose rows are indexed by $|i\rangle$, columns by $|j\rangle$, and $b_{ij} = 1$ if $i \cap j = \emptyset$, -1 if $i \cap j \neq \emptyset$. Since M_χ is a linear combination of B and the ones matrix, B has the same eigenvectors as M_χ .

B is real and symmetric, and therefore has an orthonormal set of eigenvectors $W = \{w_1, \dots, w_{N-1}\}$. Using a result of Lovasz for B [?], we characterize the eigenspaces of B as follows:

Lemma 2 [?] *B has $k+1$ eigenspaces. Eigenspace E_0 is of dimension 1 and contains the all 1's vector. E_i has dimension $n()i - n()i - 1$. The typical eigenvector in E_i is indexed by $x_1, \dots, x_{2i-1}, x_{2i} \in \{1, \dots, n\}$. The corresponding eigenvector e (unnormalized) is given by: $e_S = 0$ if there is an index $j : |S \cap \{x_{2j-1}, x_{2j}\}| \neq 1$, otherwise $e_S = \prod_j (-1)^{|S \cap x_{2j}|}$. The corresponding eigenvalues are $\lambda_0 = \frac{2n-k()k-n()k}{n()k}$, and $\lambda_i = \frac{2n-k-i()k-i}{n()k}$.*

The first g eigenspaces contain less than $gn()g \leq n^{g+1}$ eigenvectors. Setting $t = n^{g+1}$, we wish to show that we can pick g small enough to satisfy $\sum_{i>t}^{|V|} \lambda_i^2 < 2\epsilon$.

A projection of a row of B onto subspace E_i is the sum of the row of B onto each vector w spanning E_i . Let q_i be the square of the length of the projection of a row b onto the subspace E_i . Since the value of such a projection is λ_i , then $q_i = \sum_{w \in E_i} \lambda_i^2$. Therefore,

$$\sum_{i>t}^{|V|} \lambda_i^2 \leq \sum_{i>g}^{k+1} \sum_{w \in E_i} \lambda_i^2 = \sum_{i>g}^k q_i.$$

Since each E_i has dimension $n()i - n()i - 1$, E_i has $n()i - n()i - 1$ eigenvectors, each with value $\lambda_i =$

$\frac{2n-k-i()k-i}{n()k}$. Therefore,

$$q_i = (n()i - n()i - 1) \left(\frac{2n-k-i()k-i}{n()k} \right)^2.$$

Next, we show that the projections decay rapidly, by showing the relative change between q_i and q_{i+1} .

$$\begin{aligned}
\frac{q_{i+1}}{q_i} &= \frac{(n()i+1 - n()i-1) \left(\frac{2n-k-(i+1)()k-(i+1)}{n()k} \right)^2}{(n()i - n()i - 1) \left(\frac{2n-k-i()k-i}{n()k} \right)^2} \\
&\leq \frac{2n}{i+1} \cdot \frac{(k-i)^2}{(n-k-i)^2} \\
&\leq \frac{2n}{i+1} \cdot \left(\frac{2k}{n} \right)^2
\end{aligned}$$

If $k = \Theta(\sqrt{n})$, then $\frac{q_{i+1}}{q_i} = O(\frac{1}{i+1})$. Therefore, $q_g \leq \frac{c^g}{g}$.

Choosing g to be $k \log n \log(1/\epsilon) / \log \log(1/\epsilon)$,

$$\sum_{i>g}^k q_i = \sum_{i>g}^k c^i / i < c^g / g! < \epsilon.$$

Since $\sum_{i>t}^{|V|} \lambda_i^2 \leq \sum_{i>g}^{k+1} q_i$,

$$\sum_{i>t}^{|V|} \lambda_i^2 \leq \epsilon.$$

With $g = O(\log(1/\epsilon) / \log \log(1/\epsilon))$, the total number of eigenvectors needed is bounded above by $n^{g+1} = O(n^{\log 1/\epsilon})$, and therefore only $O(\log n \log 1/\epsilon)$ entangled qubits are needed to perform the protocol. This completes the proof. \blacksquare

9. Acknowledgements

The author wishes to thank Umesh Vazirani, Richard Josza, and Jennifer Sokol for numerous helpful discussions. This research is supported by a Siebel Scholars Fellowship and DARPA Grant F30602-00-2-060.

A. Proof of Remark 1

We prove that given a bipartite state $|\psi\rangle$ and local unitary operations U and V , we can write the result of $(U \otimes V) |\psi\rangle$ as UMV^\dagger where M is the matrix representing $|\psi\rangle$.

Proof:

A and B want to perform $U \otimes V$ on a bipartite state $|\phi\rangle = \sum_{ij} \phi_{ij} |i\rangle |j\rangle$. The coefficient ϕ_{ij} refers to the

probability amplitude that a measurement of the state $|\phi\rangle$ will yield $|i\rangle$ to party A and $|j\rangle$ to party B . Performing the unitary operation $U \otimes V$ on $|\phi\rangle$ changes the coefficients to $|\phi'\rangle$ according to the rule

$$\begin{aligned}\phi'_{ij} &= \langle (i\text{th row of } U \otimes j\text{th row of } V | \phi \rangle \\ &= \sum_k \sum_m U_{ik} V_{jm} \phi_{km} \\ &= \sum_k U_{ik} \sum_m V_{jm} \phi_{km} \\ &= \sum_k U_{ik} \sum_m \phi_{km} V_{jm}\end{aligned}$$

Now, we show that in the matrix formulation, $U \otimes V|\phi\rangle = U M V^\dagger$. Beginning with a matrix M , we show the values of the matrix $U M V^\dagger$.

$$\begin{aligned}(UM)_{ij} &= \sum_{k=1}^n U_{ik} M_{kj} \\ (AV^T)_{ij} &= \sum_{m=1}^n A_{im} V_{jm} \\ (UAV^T)_{ij} &= \sum_k U_{ik} \left(\sum_m A_{km} V_{jm} \right) \\ (UMV^T)_{ij} &= \sum_k U_{ik} \left(\sum_m M_{km} V_{jm} \right)\end{aligned}$$

Hence, these two formulations are equal. \blacksquare

B. Proof of Remark 2

Proof:

Note that $\text{Tr}(M_\chi^\dagger M_\chi) = \langle \chi | \chi \rangle$, since $\text{Tr}(M_\chi)^\dagger M_\chi = \sum_{ij} (M_\chi)_{ij}$. But since this matrix consists of all the elements of $|\chi\rangle$, this result is $\sum_{ij} (M_{\chi ij}) = \langle \chi | \chi \rangle$.

We rewrite $\|VM_\psi V^\dagger - M_\chi\|^2$ in terms of the trace norm:

$$\begin{aligned}\|VM_\psi V^\dagger - M_\chi\|^2 &= \\ &\text{Tr} \left((VM_\psi V^\dagger)^\dagger (VM_\psi V^\dagger) \right) - \\ &2\text{Tr} (M_\chi^\dagger VM_\psi V^\dagger) + \text{Tr} (M_\chi^\dagger M_\chi)\end{aligned}$$

Since the trace is invariant under basis transformations, and $V^\dagger V = I$,

$$\begin{aligned}\text{Tr} ((VM_\psi V^\dagger)^\dagger (VM_\psi V^\dagger)) &= \text{Tr} ((VM_\psi^\dagger V^\dagger VM_\psi V^\dagger)) \\ &= \text{Tr} ((M_\psi^\dagger M_\psi)) \\ &= \sum_{ij} (M_{\psi ij}) = \langle \psi | \psi \rangle\end{aligned}$$

Continuing, we see that

$$\begin{aligned}-2\text{Tr} (VM_\psi^\dagger V^\dagger M_\chi) &= -2\text{Tr} (VM_\psi^\dagger V^\dagger VAV^\dagger) \\ &= -2\text{Tr} (M_\psi^\dagger \Lambda) \\ &= -2\langle \psi, 000 | \chi \rangle\end{aligned}$$

where $\langle 000 |$ represents an ancilla, so that the $|\chi\rangle$ and $|\psi, 000\rangle$ lie in the same vector space.

Therefore,

$$\begin{aligned}\|VM_\psi V^\dagger - M_\chi\|^2 &= \langle \psi | \psi \rangle + \langle \chi | \chi \rangle - 2\langle \psi, 000 | \chi \rangle \\ &= 2 - 2\langle \psi, 000 | \chi \rangle \\ &< 2\epsilon\end{aligned}$$

Therefore, $\langle \chi | \psi, 000 \rangle > 1 - \epsilon$. Since unitary operations on pure states cannot change their length, $\langle \chi | \psi, 000 \rangle = \langle \chi | V \otimes V | \psi, 000 \rangle > 1 - \epsilon$. \blacksquare